A
[30]
図3

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号
特開平9－105748
(43)公開日　平成９年(1997)４月22日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０１Ｎ　33/50			Ｇ０１Ｎ　33/50	Ｐ	
Ｇ０６Ｆ　17/30			Ｇ０６Ｆ　15/403	３５０Ａ	
// Ｃ１２Ｎ　15/09		9162－4B	Ｃ１２Ｎ　15/00	Ａ	

審査請求　未請求　請求項の数10　ＯＬ　（全 10 頁）

(21)出願番号　特願平7－265157

(22)出願日　平成７年(1995)10月13日

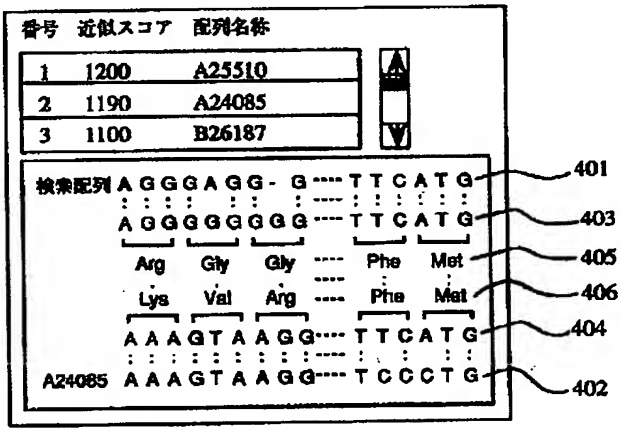
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　平岡　進
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　永井　啓一
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　西川　哲夫
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内
(74)代理人　弁理士　小川　勝男

最終頁に続く

(54)【発明の名称】　ＤＮＡ配列比較方法

(57)【要約】
【課題】翻訳されるアミノ酸配列が互いに類似するＤＮＡ配列を検索することができるＤＮＡ配列比較方法を提供する。
【解決手段】検索ＤＮＡ配列とデータベース中の既知ＤＮＡ配列に対して、３文字毎にアミノ酸に翻訳ための中間ＤＮＡ配列を作成し、中間ＤＮＡ配列を翻訳してアミノ酸配列を作成する。検索ＤＮＡ配列を翻訳するための中間ＤＮＡ配列と検索ＤＮＡ配列を対応させて表示し、既知ＤＮＡ配列を翻訳するための中間ＤＮＡ配列と既知ＤＮＡ配列を対応させて表示する。中間ＤＮＡ配列の３文字とアミノ酸を対応させてアミノ酸配列を表示する。対応するアミノ酸配列の間、検索ＤＮＡ配列と対応する中間ＤＮＡ配列の間、および既知ＤＮＡ配列と対応する中間ＤＮＡ配列の間に類似度を示す記号列を表示する。

図3



1

【特許請求の範囲】

【請求項１】第一と第二のＤＮＡ配列の塩基配列を比較するＤＮＡ配列比較方法において、前記第一と第二のＤＮＡ配列を所定の長さの塩基群に分割し、分割された塩基群をアミノ酸に翻訳し、前記第一のＤＮＡ配列から翻訳されたアミノ酸を第一のアミノ酸配列として、前記第二のＤＮＡ配列から翻訳されたアミノ酸を第二のアミノ酸配列として、前記第一のアミノ酸配列と前記第二のアミノ酸配列とを並置して、前記塩基群とともに表示することを特徴とするＤＮＡ配列比較方法。

【請求項２】前記所定の長さの塩基群は、挿入、または欠失を含み、前記挿入、または前記欠失を表示することを特徴とする請求項１に記載のＤＮＡ配列比較方法。

【請求項３】前記第一と第二のＤＮＡ配列を所定の方向に塩基を順次シフトして、前記塩基群に分割し、前記第一と第二のＤＮＡ配列から翻訳された前記第一と第二のアミノ酸配列の間で、それぞれのアミノ酸について類似度を積算し、積算結果が最大になるように、前記第一と第二のＤＮＡ配列を所定の方向に塩基を順次シフトして、前記第一と第二のアミノ酸配列を選択することを特徴とする請求項１に記載のＤＮＡ配列比較方法。

【請求項４】前記第二のＤＮＡ配列が、ＤＮＡ配列データベースから選択された既知ＤＮＡ配列であることを特徴とする請求項１に記載のＤＮＡ配列比較方法。

【請求項５】所定の塩基長の固定長ＤＮＡ配列とＤＮＡ配列データベース内の既知ＤＮＡ配列とを所定の方向に塩基を順次シフトして、３塩基長からなる塩基群に分割し、前記既知ＤＮＡ配列と前記固定長ＤＮＡ配列との分割された前記塩基群のそれぞれをアミノ酸に翻訳し、前記既知ＤＮＡ配列と前記固定長ＤＮＡ配列とから翻訳されたアミノ酸からなるアミノ酸配列の間でそれぞれのアミノ酸について類似度を積算し、前記既知ＤＮＡ配列について、積算結果の最大値を前記固定長ＤＮＡ配列に対応した構成要素とするスコア表を作成し、前記第一のＤＮＡ配列に対して前記既知ＤＮＡ配列毎に、第一のＤＮＡ配列を前記所定の塩基長に分割し、前記既知ＤＮＡ配列と第一のＤＮＡ配列との所定の塩基長毎の類似度の和を前記スコア表の類似度を参照して求め、求めた類似度の和に対応させて順次前記既知ＤＮＡ配列を表示し、求めた類似度の和の高い前記既知ＤＮＡ配列を前記第二のＤＮＡ配列とすることを特徴とする請求項１に記載のＤＮＡ配列比較方法。

【請求項６】前記第一と第二のＤＮＡ配列を第一のフォントで表示し、前記第一と第二のアミノ酸配列を第二のフォントで表示することを特徴とする請求項１に記載のＤＮＡ配列比較方法。

【請求項７】第一と第二のＤＮＡ配列の塩基配列を比較するＤＮＡ配列比較方法において、前記第一のＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第一の中間ＤＮＡ配列を作成し、前記第二のＤＮＡ

2

配列から、アミノ酸に翻訳するために３塩基ずつに区分された第二の中間ＤＮＡ配列を作成し、区分された前記第一と第二の中間ＤＮＡ配列をそれぞれの３塩基ごとにアミノ酸に翻訳し、前記第一と第二の中間ＤＮＡ配列のそれぞれの３塩基ごとに対応させて翻訳された前記アミノ酸のそれぞれを並置して、前記第一の中間ＤＮＡ配列から翻訳されたアミノ酸を第一のアミノ酸配列として、前記第二の中間ＤＮＡ配列から翻訳されたアミノ酸を第二のアミノ酸配列として、表示し、前記第一のＤＮＡ配列と前記第一の中間ＤＮＡ配列の間の第一の類似度を求め、前記第二のＤＮＡ配列と前記第二の中間ＤＮＡ配列の間の第二の類似度を求め、前記第一と第二のアミノ酸配列の間の第三の類似度を求め、前記第一、第二および第三の類似度から所定の関数を用いて得られるパラメータが最大となるように前記第一と第二の中間ＤＮＡ配列および前記第一と第二のアミノ酸配列を選択することを特徴とするＤＮＡ配列比較方法。

【請求項８】前記第二のＤＮＡ配列が、ＤＮＡ配列データベースから選択された既知ＤＮＡ配列であることを特徴とする請求項７に記載のＤＮＡ配列比較方法。

【請求項９】ＤＮＡ配列データベース内の既知ＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第三の中間ＤＮＡ配列を作成し、所定の塩基長の固定長ＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第四の中間ＤＮＡ配列を作成し、前記第三と第四の中間ＤＮＡ配列の区分されたそれぞれの３塩基ごとにアミノ酸に翻訳し、前記第三と第四の中間ＤＮＡ配列のそれぞれの３塩基ごとに対応させて翻訳された前記アミノ酸のそれぞれを並置して、前記第三の中間ＤＮＡ配列から翻訳されたアミノ酸を第三のアミノ酸配列として、前記第四の中間ＤＮＡ配列から翻訳されたアミノ酸を第四のアミノ酸配列として、前記既知ＤＮＡ配列と前記第三の中間ＤＮＡ配列との間の第四の類似度を求め、前記固定長ＤＮＡ配列と前記第四の中間ＤＮＡ配列との間の第五の類似度を求め、前記第三と第四のアミノ酸配列の間の第六の類似度を求め、前記既知ＤＮＡ配列について、前記第四、第五および第六の類似度から所定の関数を用いて得られるパラメータの最大値を前記固定長ＤＮＡ配列に対応した構成要素とするスコア表を作成し、前記第一のＤＮＡ配列に対して前記既知ＤＮＡ配列毎に、第一のＤＮＡ配列を前記所定の塩基長に分割し、前記既知ＤＮＡ配列と第一のＤＮＡ配列との所定の塩基長毎の類似度の和を前記スコア表の類似度を参照して求め、求めた類似度の和に対応させて順次前記既知ＤＮＡ配列を表示し、求めた類似度の和の高い前記既知ＤＮＡ配列を前記第二のＤＮＡ配列とすることを特徴とする請求項７に記載のＤＮＡ配列比較方法。

【請求項１０】前記第一と第二のＤＮＡ配列を第一のフォントで表示し、前記第一と第二のアミノ酸配列を第二のフォントで表示することを特徴とする請求項７に記載

のＤＮＡ配列比較方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、ＤＮＡ配列比較方法に関し、特にＤＮＡ配列の照合に適したＤＮＡ配列比較方法に関する。

【０００２】

【従来の技術】近年、遺伝子工学の発展によりＤＮＡ配列に関するデータが急増している。ＤＮＡは生物の設計図と言うことができる物質であり、生物の機能は全てＤＮＡ中に記述されている。ＤＮＡはＡ、Ｃ、Ｇ、Ｔの文字で代表される４種類の物質が並んだ構造となっており、Ａ、Ｃ、Ｇ、Ｔだけの文字列であるＤＮＡ配列として表すことができる。

【０００３】生体中ではＤＮＡの一部が翻訳されて蛋白となり生物機能を担っている。重要な生物機能を担っている蛋白を見つけることが出来れば薬として用いることも可能である。インターフェロン、インスリンなどは薬として製造されるようにもなった蛋白の例である。蛋白構造全体がわからない場合でも薬物などと結合する結合サイトがわかれば、その蛋白が関係する生物機能を制御する薬の基本構造を決定することができる。蛋白は２０種類のアミノ酸が一列に並んだ構造となっており、ＤＮＡと同様にその構造はアミノ酸配列として文字列で表すことができる。ＤＮＡ配列からアミノ酸配列への翻訳はコドンと呼ばれるＤＮＡの３文字が一組となり行われる。図６に示すようにコドンからアミノ酸への翻訳規則は解析されており、コドンからアミノ酸への翻訳は一意的に行われる。ＤＮＡ配列の翻訳方向と翻訳開始点がわかればＤＮＡ配列から３文字ずつコドンを取り出すことで生物中で行われているのと同様にＤＮＡから蛋白の翻訳を行うことができる。ただしＤＮＡ配列中に挿入、欠失などの測定誤りがあった場合には、途中からコドンの枠がずれてしまうため蛋白への翻訳を誤ってしまう。

【０００４】ＤＮＡ配列の翻訳方向と翻訳開始点を示す規則は完全にはわかっていない。そのためコドンからアミノ酸への翻訳と異なりＤＮＡ配列中の翻訳領域の決定は容易には行われていない。またコドン枠がずれないようにほとんど誤りなくＤＮＡ配列を決定するためには配列決定作業が増加してしまう。そのためＤＮＡ配列が決定されても対応する蛋白配列の決定が容易に行われないことも多く、未知のＤＮＡ配列から未知の蛋白を発見することは容易ではない。未知の蛋白の発見は新薬の開発に重要である。

【０００５】ＤＮＡ配列から翻訳領域を決定するなどして蛋白を決定する手法には、翻訳領域推定ソフト、蛋白測定、ｃＤＮＡ測定、ハイブリダイゼーション、類似ＤＮＡ配列検索などがある。翻訳領域推定ソフトはこれまでわかっている知識を総合してＤＮＡ配列から蛋白翻訳領域を推定するソフトである。蛋白翻訳に関する知識が

完全ではないため精度はあまり高くない。蛋白測定は生物中で合成されている蛋白を直接測定する方法であり、感度が高くなくアミノ酸配列測定も容易でないため、大量の生物試料が必要とされる。

【０００６】ｃＤＮＡ測定はＤＮＡから蛋白へ翻訳される過程の中間物質をＤＮＡに変換して測定する方法である。蛋白と異なりＤＮＡは増幅可能なため、ｃＤＮＡ測定は蛋白測定に比べて感度は高い。それでも微量な蛋白に対応するｃＤＮＡは大量な蛋白に対応するｃＤＮＡの中に混合し埋もれてしまうため、特殊な蛋白を測定するためには特殊な生物試料からｃＤＮＡを抽出するか、ハイブリダイゼーション、ＤＮＡ配列相同性検索などを用いて特定のｃＤＮＡを選択する必要がある。

【０００７】ハイブリダイゼーションはすでに手元にあるＤＮＡに対して類似しているＤＮＡをＤＮＡ混合試料または長いＤＮＡから選択する技術である。手元にあるＤＮＡに相補的なＤＮＡを合成し、これに結合するＤＮＡを選択することで類似ＤＮＡを選択することができる。

【０００８】ＤＮＡ配列相同性検索（ホモロジーサーチ）はハイブリダイゼーションが生物学的手法で行っていることを情報処理的手法で行うものである。あらかじめ全てのＤＮＡ配列を測定しておく必要があるが、ハイブリダイゼーションと異なり相補的結合を行うようなＤＮＡばかりでなく、異なった基準で類似しているＤＮＡ配列を選択することができる。

【０００９】相同性検索を行う最も基本的な方法として、問い合わせ配列とデータベース内の各配列との間でスミスとウォーターマンにより提案されたダイナミックプログラミング法によるアライメントを行い、高いスコア順に各配列を表示する方法がある。ダイナミックプログラミング法によるアライメントに関してはジャーナル・オブ・モレキュラー・バイオロジー、１４７（１９８１年）１９５頁から１９７頁（Ｊｏｕｎａｌ　ｏｆ　Ｍｏｌｅｃｕｌａｒ　Ｂｉｏｌｏｇｙ、１４７（１９８１）ｐｐ１９５－１９７）に記載されている。

【００１０】ダイナミックプログラミング法による配列１（ＣＡＧＴＧＡＣＴ）と配列２（ＣＡＣＴＧＣＴＧ）のアライメントを図７を用いて説明する。ダイナミックプログラミング法によるアライメントでは２次元メッシュのＸ、Ｙ方向に添ってそれぞれ２本の配列を置き、メッシュの各点をノードとして、ノード間には縦、横、斜めの３方向の経路を考えた時に任意の２つのノード間を左上から右下に向かう最適経路を求める。縦、横のアーク（格子点間を結ぶ線）は挿入・欠失に相当するためペナルティスコアがかかり、また配列要素が対合する斜めのアークにも対合の種類に応じたスコアが与えられる。これらのスコアを経路に添って総計した合計スコアがもっとも高くなる経路をダイナミックプログラミング法によって解き最適なアライメントを求める。ＤＮＡ配列と

5

うしのアライメントにおいて一般的に用いられているスコアは、挿入・欠失のスコアはｎ文字の挿入・欠失に対して－４ｎ－８点、一致した１文字のスコアは４点、異なっている1文字のスコアは－３点である。例えば図７に示した経路でのスコアは９点である。

【００１１】相同性検索プログラムであるＦＡＳＴＡについては、アカデミックプレス（Ａｃａｄｅｍｉｃ　Ｐｒｅｓｓ）より発行されたドゥーリトル（Ｄｏｏｌｉｔｔｌｅ）編集によるメソッヅ・イン・エンザイモロジー、１８３（１９９０年）第６３頁から９８頁（Ｍｅｔｈｏｄｓ　ｉｎ　Ｅｎｚｙｍｏｌｏｇｙ、１８３（１９９０）ｐｐ６３－９８）に記載されている。ＦＡＳＴＡではダイナミックプログラミング法によるスコアよりも少ない計算量で求められるｉｎｉｔｎ、ｉｎｉｔ１と呼ばれるスコアを求めている。これらのスコアは完全に一致する一定長の部分文字列を探し出し、それらを繋ぎあわせて求める。ＦＡＳＴＡではこれらのスコアが高い順に配列を表示しており、さらに上位の配列に対してはダイナミックプログラミング法によるスコアを計算しＯＰＴという名で表示している。

【００１２】また、相同性検索プログラムであるＢＬＡＳＴについては、ジャーナル・オブ・モレキュラー・バイオロジー、２１５（１９９０年）第４０３頁から４１０頁（Ｊｏｕｎａｌ　ｏｆ　Ｍｏｌｅｃｕｌａｒ　Ｂｉｏｌｏｇｙ、２１５（１９９０）ｐｐ４０３－４１０）に記載されている。ＢＬＡＳＴでもダイナミックプログラミング法によるスコアよりも少ない計算量で求められるスコアを求めている。ＢＬＡＳＴではＦＡＳＴＡにおけるｉｎｉｔ１よりも単純化したスコアを用いている。ＢＬＡＳＴではＢＬＡＳＴ独自のスコア計算のみであり、ダイナミックプログラミング法によるスコアの計算は行っていない。

【００１３】スミスとウォーターマンによるダイナミックプログラミング法、ＦＡＳＴＡ、ＢＬＡＳＴはいずれもＤＮＡ配列をＤＮＡ配列データベースに対して検索する方法であり、ＤＮＡ配列どうしを比較している。これらのプログラムはアミノ酸配列どうしを比較することで、アミノ酸配列をアミノ酸配列データベースに対して検索することも可能である。アミノ酸配列の比較で用いられているスコアを図８に示す。

【００１４】これに対してＤＮＡ配列をアミノ酸配列データベースに対して検索するＢＬＡＳＴＸ、そしてアミノ酸配列をＤＮＡ配列データベースに対して検索するＴＦＡＳＴＡ、ＴＢＬＡＳＴＮが存在する。これらは比較する際にＤＮＡ配列を方向と翻訳開始点を変えた６通りのアミノ酸配列に翻訳し、アミノ酸配列どうしで比較している。

【００１５】

【発明が解決しようとする課題】従来の配列比較、検索プログラムでは、ＤＮＡ配列とアミノ酸配列の比較において

6

ＤＮＡ配列の翻訳を行っていたが、ＤＮＡ配列どうしの比較においては両ＤＮＡ配列をアミノ酸に翻訳し比較することを行っていなかった。ＤＮＡ配列のアミノ酸翻訳領域はたとえ翻訳されたアミノ酸がわからない場合でも、ＤＮＡ配列そのままで比較するよりもアミノ酸配列に翻訳して比較することが望ましい。類似機能を持つ蛋白は類似アミノ酸配列部分を持ち、その部分が共通する重要な働きをしているからである。

【００１６】コドンからアミノ酸への翻訳は多対一であるため異なるＤＮＡ配列でも同じ蛋白に翻訳されることがある。さらにコドンが完全に異なり異なるアミノ酸に翻訳される場合でも、それらのアミノ酸の性質が類似していることがある。言い替えれば翻訳されたアミノ酸配列が類似していても元のＤＮＡ配列が類似しているとは限らない。アミノ酸配列に翻訳して比較することで翻訳されたアミノ酸配列が類似しているＤＮＡ配列と類似している部分を選択することができる。類似アミノ酸配列に翻訳される部分はアミノ酸配列翻訳領域であり、蛋白の共通機能部分であることが推定できる。

【００１７】従来のＤＮＡ配列検索ではＤＮＡ配列そのものが類似している順番にＤＮＡ配列データベースを並べていた。そのため翻訳されたアミノ酸が類似しているＤＮＡ配列は多くの偶然似ているＤＮＡ配列に埋もれてしまって見つけられない点に問題がある。ＤＮＡ配列どうしを一対一で比較する場合にも多くの偶然類似しているＤＮＡ部分配列の中に埋もれてしまい、翻訳されたアミノ酸配列が類似しているＤＮＡ配列部分を見つけることは困難である点に問題がある。

【００１８】本発明の目的は、ＤＮＡ配列から翻訳されたアミノ酸配列を比較することで、翻訳されるアミノ酸配列が互いに類似するＤＮＡ配列をもれなく検索することができるＤＮＡ配列比較方法を提供することにある。

【００１９】

【課題を解決するための手段】上記目的を達成するために本発明では、比較するＤＮＡ配列をアミノ酸配列に翻訳し、翻訳されたアミノ酸配列を比較することを特徴とする。翻訳したアミノ酸配列の類似部分配列を対応させてアミノ酸配列を並置表示し、ＤＮＡ配列の各コドンを翻訳された各アミノ酸に対応させて表示する。ＤＮＡ配列からアミノ酸配列への翻訳と、アミノ酸配列どうしの比較結果を表示する。ＤＮＡ配列をアミノ酸配列に翻訳し、翻訳されたアミノ酸を比較することで、翻訳されたアミノ酸配列が類似しているＤＮＡ配列を見つけるこてとができ、翻訳されたアミノ酸配列が互いに類似するＤＮＡ配列をもれなく検索することができる。

【００２０】以下、より詳細に本発明の特徴を説明すると、第一と第二のＤＮＡ配列を所定の長さの塩基群に分割し、分割された塩基群をアミノ酸に翻訳し、第一のＤＮＡ配列から翻訳されたアミノ酸を第一のアミノ酸配列として、第二のＤＮＡ配列から翻訳されたアミノ酸を第

二のアミノ酸配列として、第一のアミノ酸配列と第二のアミノ酸配列とを並置して、塩基群とともに表示することを特徴とする。所定の長さの塩基群は、挿入、または欠失を含み、挿入、または欠失を表示する。第一と第二のＤＮＡ配列を所定の方向に塩基を順次シフトして、塩基群に分割し、第一と第二のＤＮＡ配列から翻訳された第一と第二のアミノ酸配列の間で、それぞれのアミノ酸について類似度を積算し、積算結果が最大になるように、第一と第二のＤＮＡ配列を所定の方向に塩基を順次シフトして、第一と第二のアミノ酸配列を選択する。第二のＤＮＡ配列が、ＤＮＡ配列データベースから選択された既知ＤＮＡ配列であってもよい。

【００２１】所定の塩基長の固定長ＤＮＡ配列とＤＮＡ配列データベース内の既知ＤＮＡ配列とを所定の方向に塩基を順次シフトして、３塩基長からなる塩基群に分割し、既知ＤＮＡ配列と固定長ＤＮＡ配列との分割された塩基群のそれぞれをアミノ酸に翻訳し、既知ＤＮＡ配列と固定長ＤＮＡ配列とから翻訳されたアミノ酸からなるアミノ酸配列の間でそれぞれのアミノ酸について類似度を積算し、既知ＤＮＡ配列について、積算結果の最大値を固定長ＤＮＡ配列に対応した構成要素とするスコア表を作成し、第一のＤＮＡ配列に対して既知ＤＮＡ配列毎に、第一のＤＮＡ配列を所定の塩基長に分割し、既知ＤＮＡ配列と第一のＤＮＡ配列との所定の塩基長毎の類似度の和をスコア表の類似度を参照して求め、求めた類似度の和に対応させて順次既知ＤＮＡ配列を表示し、求めた類似度の和の高い既知ＤＮＡ配列を第二のＤＮＡ配列としてもよい。第一と第二のＤＮＡ配列を第一のフォントで表示し、第一と第二のアミノ酸配列を第二のフォントで表示する。

【００２２】また、本発明では、第一のＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第一の中間ＤＮＡ配列を作成し、第二のＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第二の中間ＤＮＡ配列を作成し、区分された第一と第二の中間ＤＮＡ配列をそれぞれの３塩基ごとにアミノ酸に翻訳し、第一と第二の中間ＤＮＡ配列のそれぞれの３塩基ごとに対応させて翻訳されたアミノ酸のそれぞれを並置して、第一の中間ＤＮＡ配列から翻訳されたアミノ酸を第一のアミノ酸配列として、第二の中間ＤＮＡ配列から翻訳されたアミノ酸を第二のアミノ酸配列として、表示し、第一のＤＮＡ配列と第一の中間ＤＮＡ配列の間の第一の類似度を求め、第二のＤＮＡ配列と第二の中間ＤＮＡ配列の間の第二の類似度を求め、第一と第二のアミノ酸配列の間の第三の類似度を求め、第一、第二および第三の類似度から所定の関数を用いて得られるパラメータが最大となるように第一と第二の中間ＤＮＡ配列および第一と第二のアミノ酸配列を選択することを特徴とする。第二のＤＮＡ配列が、ＤＮＡ配列データベースから選択された既知ＤＮＡ配列であってもよい。

【００２３】ＤＮＡ配列データベース内の既知ＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第三の中間ＤＮＡ配列を作成し、所定の塩基長の固定長ＤＮＡ配列から、アミノ酸に翻訳するために３塩基ずつに区分された第四の中間ＤＮＡ配列を作成し、第三と第四の中間ＤＮＡ配列の区分されたそれぞれの３塩基ごとにアミノ酸に翻訳し、第三と第四の中間ＤＮＡ配列のそれぞれの３塩基ごとに対応させて翻訳されたアミノ酸のそれぞれを並置して、第三の中間ＤＮＡ配列から翻訳されたアミノ酸を第三のアミノ酸配列として、第四の中間ＤＮＡ配列から翻訳されたアミノ酸を第四のアミノ酸配列として、既知ＤＮＡ配列と第三の中間ＤＮＡ配列との間の第四の類似度を求め、固定長ＤＮＡ配列と第四の中間ＤＮＡ配列との間の第五の類似度を求め、第三と第四のアミノ酸配列の間の第六の類似度を求め、既知ＤＮＡ配列について、第四、第五および第六の類似度から所定の関数を用いて得られるパラメータの最大値を固定長ＤＮＡ配列に対応した構成要素とするスコア表を作成し、第一のＤＮＡ配列に対して既知ＤＮＡ配列毎に、第一のＤＮＡ配列を所定の塩基長に分割し、既知ＤＮＡ配列と第一のＤＮＡ配列との所定の塩基長毎の類似度の和をスコア表の類似度を参照して求め、求めた類似度の和に対応させて順次既知ＤＮＡ配列を表示し、求めた類似度の和の高い既知ＤＮＡ配列を第二のＤＮＡ配列としてもよい。第一と第二のＤＮＡ配列を第一のフォントで表示し、第一と第二のアミノ酸配列を第二のフォントで表示する。

【００２４】

【発明の実施の形態】本発明を検索ＤＮＡ配列に対するＤＮＡ配列データベース検索に応用した実施例について詳述する。あらかじめＤＮＡ配列データベースに対してスコア表を作成しておく。スコア表はＤＮＡ配列データベース内の各ＤＮＡ配列について、所定長の固定長ＤＮＡ配列に対応したアミノ酸配列に翻訳した時のスコア（類似度）を、構成要素とする表である。図１を用いてＤＮＡ配列データベース内のＤＮＡ配列と固定長ＤＮＡ配列とをアミノ酸配列に翻訳した時のスコアの計算方法を説明する。

【００２５】ＤＮＡ配列データベース内のＤＮＡ配列をＤＮＡ配列３０１、固定長ＤＮＡ配列をＤＮＡ配列３０２とする。ＤＮＡ配列３０１から中間ＤＮＡ配列３０３を生成し、ＤＮＡ配列３０２から中間ＤＮＡ配列３０４を生成する。中間ＤＮＡ配列３０３、３０４はアミノ酸配列に翻訳するために３塩基ずつに区分されている。３塩基ずつ翻訳していくことで、中間ＤＮＡ配列３０３からアミノ酸配列３０５が生成され、中間ＤＮＡ配列３０４からアミノ酸配列３０６が生成される。

【００２６】ＤＮＡ配列３０１と中間ＤＮＡ配列３０３をダイナミックプログラミング法を用いてアライメント（並置）し、スコア３０７を求め、ＤＮＡ配列３０２と

中間ＤＮＡ配列３０４をダイナミックプログラミング法を用いてアライメントし、スコア３０８を求め、アミノ酸配列３０５とアミノ酸配列３０６をダイナミックプログラミング法を用いてアライメントし、スコア３０９を求める。スコア３０７、３０８、３０９は用いる中間ＤＮＡ配列３０３、３０４によって様々な値をとる。適当な中間ＤＮＡ配列３０３、３０４を用いて、スコア３０７、３０８、３０９から所定の関数を用いて得られるパラメータの最大値を求める。この最大値がＤＮＡ配列３０１とＤＮＡ配列３０２をアミノ酸配列に翻訳したときの類似性を示すスコアとなる。この値をスコア表に格納する。

【００２７】ＤＮＡ配列データベース検索の際は、検索ＤＮＡ配列を固定長ＤＮＡ配列と同じ塩基長に分割し、ＤＮＡ配列データベース内の各ＤＮＡ配列と検索ＤＮＡ配列との固定長ＤＮＡ配列の塩基長毎のスコアの和をスコア表のスコアを参照して求め近似スコアとする。前記近似スコアに対応させて順次ＤＮＡ配列データベース内の各ＤＮＡ配列の配列名称を近似スコアの値の大きい順に近似スコアと共に表示する。

【００２８】表示されたＤＮＡ配列データベース内の各ＤＮＡ配列の中からＤＮＡ配列を選択し、選択されたＤＮＡ配列と検索ＤＮＡ配列とをアミノ酸に翻訳したときの類似性を示すスコアを求める計算を行う。このスコアの求め方を図２を用いて以下で説明する。

【００２９】ＤＮＡ配列データベース内のＤＮＡ配列をＤＮＡ配列４０２、検索ＤＮＡ配列をＤＮＡ配列４０１とする。ＤＮＡ配列４０１から中間ＤＮＡ配列４０３を生成し、ＤＮＡ配列４０２から中間ＤＮＡ配列４０４を生成する。中間ＤＮＡ配列４０３、４０４はアミノ酸配列に翻訳するために３塩基ずつに区分されている。３塩基ずつ翻訳していくことで、中間ＤＮＡ配列４０３からアミノ酸配列４０５が生成され、中間ＤＮＡ配列４０４からアミノ酸配列４０６が生成される。ＤＮＡ配列４０１と中間ＤＮＡ配列４０３をダイナミックプログラミング法を用いてアライメント（並置）し、スコア４０７を求め、ＤＮＡ配列４０２と中間ＤＮＡ配列４０４をダイナミックプログラミング法を用いてアライメントし、スコア４０８を求め、アミノ酸配列４０５とアミノ酸配列４０６をダイナミックプログラミング法を用いてアライメントし、スコア４０９を求める。スコア４０７、４０８、４０９は用いる中間ＤＮＡ配列４０３、４０４によって様々な値をとる。中間ＤＮＡ配列４０３、４０４には、スコア４０７、４０８、４０９から所定の関数を用いて得られるパラメータが最大値となるＤＮＡ配列を選択する。パラメータの最大値はＤＮＡ配列４０１とＤＮＡ配列４０２をアミノ酸配列に翻訳したときの類似性を示すスコアである。

【００３０】本実施例を用いた検索結果を図３に示す。近似スコアに対応させて順次ＤＮＡ配列データベース内の各ＤＮＡ配列の配列名称を近似スコアの値の大きい順に近似スコアと共に表示する。ＤＮＡ配列４０１（検索ＤＮＡ配列）と中間ＤＮＡ配列４０３を対応させて表示し、ＤＮＡ配列４０２（ＤＮＡ配列データベース内のＤＮＡ配列（Ａ２４０８５））と中間ＤＮＡ配列４０４とを対応させて表示し、アミノ酸配列４０５とアミノ酸配列４０６を対応させて表示し、対応するＤＮＡの塩基が一致していれば記号”：”、類似していれば記号”．”、一致も類似もしていなければ記号としてブランクを対応するＤＮＡの間に表示し、対応するアミノ酸が一致していれば記号”：”、類似していれば記号”．”、一致も類似もしていなければ記号としてブランクを対応するアミノ酸の間に表示する。

【００３１】スコア表に用いる固定長配列の長さは長い方が望ましい。固定長ＤＮＡ配列の長さが短い場合には、スコア表のほとんど全ての構成要素がおなじ値になり、スコア表から計算される近似スコアもほとんど同じ値になり、結果として検索ＤＮＡ配列と翻訳したアミノ酸が似ているデータベース内の配列を選ぶことが出来ないからである。例えば固定長ＤＮＡ配列の長さとして６文字を用いた場合には、翻訳されるアミノ酸配列の長さは２文字である。データベース中のＤＮＡ配列を翻訳すると、かなりの頻度で固定長ＤＮＡ配列から翻訳されたアミノ酸とデータベース中のＤＮＡ配列から翻訳されたアミノ酸が同じになることが起こり得る。これによりスコア表のほとんどの構成要素の値は完全に一致した場合のスコアとなる。そこでスコア表に用いる固定長配列の長さとしては７文字以上が望ましい。

【００３２】スコアの関数としては最も単純には３つのスコアの和が考えられる。単純な和以外にも自乗和なども可能である。そしてＤＮＡ配列どうしのスコアとアミノ酸配列どうしのスコアも、従来用いられているスコアそのものに限定する必要はない。例えばＤＮＡどうしのスコアを半分にすれば、より大きいＤＮＡ配列測定誤差を許容した比較が可能となる。ＤＮＡ配列決定方法によってＤＮＡ配列測定誤差は異なるため、それに応じてスコアの関数、ＤＮＡ配列どうしのスコア、そしてアミノ酸配列どうしのスコアを使い分けることが望ましい。

【００３３】中間ＤＮＡ配列４０３と中間ＤＮＡ配列４０４は計算時間が問題とならない場合には、あらゆるＤＮＡ配列を生成することが望ましい。この場合、ＤＮＡ配列どうしの直接比較における場合と同様にダイナミックプログラミング法によるスコア計算が可能である。

【００３４】図４を参照して、中間ＤＮＡ配列を用いてアミノ酸に翻訳し比較するときのダイナミックプログラミング法によるスコア計算方法を以下に示す。２次元メッシュのＹ方向に沿ってＤＮＡ配列５０１を置き、Ｘ方向に沿ってＤＮＡ配列５０２を置いた。図４ではメッシュの各点がノード１０１～１１３を示し、格子間を結ぶ線分がアーク２０１～２０６を示す。各ノードを左上か

ら右下につないでいった経路がアライメント（並置）に対応しており、アークは経路を分割したときの最小単位である。そして各ノードにおけるスコアはＤＮＡ配列５０１とＤＮＡ配列５０２のそこまでの部分配列どうしを比較したときのスコアに対応している。

【００３５】ダイナミックプログラミング法では、メッシュ中のあらゆる経路のスコアを求め最大値を決定する計算を、左上から右下に向かって各ノードのスコアを順番に求めていく計算に置き換えて計算の高速化を果たしている。各ノードのスコアはそのノードで終る経路のスコアの最大値を示している。各ノードのスコアは、そのノードと単独のアークで結ばれる上、左または左上の各ノードのスコアとそこからのアークのスコアを加えたスコアを求め、それらの最大値を求めることで行われる。ダイナミックプログラミング法では全てのノードのスコアを求めた後、全てのノードのスコアの中の最大値を求めることでＤＮＡ配列５０１とＤＮＡ配列５０２の間のスコアを決定することができる。

【００３６】ＤＮＡ配列を直接比較する従来のダイナミックプログラミング法では、ノード１０１で終るアークは、ＤＮＡ挿入に対応するアーク２０１、ＤＮＡ欠失に対応するアーク２０２、ＤＮＡどうしの比較に対応するアーク２０３である。ノード１０２のスコアとアーク２０１のスコアの和、ノード１０５のスコアとアーク２０２のスコアの和、ノード１０９のスコアとアーク２０３のスコアの和、および０の中の最大値がノード１０１におけるスコアとなる。

【００３７】ＤＮＡ配列を中間ＤＮＡ配列を用いてアミノ酸配列に翻訳して比較するときのダイナミックプログラミング法では、従来のＤＮＡ配列を直接比較するときのダイナミックプログラミング法とは異なったアークを用いる必要がある。ノード１０１に達するアークには、コドンの境界へのＤＮＡ挿入に対応するアーク２０１、コドンの境界へのＤＮＡ欠失に対応するアーク２０２、翻訳したアミノ酸の挿入に対応するアーク２０４、翻訳したアミノ酸の欠失に対応するアーク２０５、翻訳したアミノ酸どうしの比較に対応するアーク２０６がある。ＤＮＡ配列の直接比較におけるダイナミックプログラミング法と異なりＤＮＡどうしの比較に対応するアーク２０３は用いない。そしてアミノ酸への翻訳では中間ＤＮＡ配列中のコドンを考慮することで、コドン中のＤＮＡの置換、挿入、欠失を含めて行う。

【００３８】例えば、アーク２０６はＤＮＡ配列５０１中のＤＮＡ３文字ＡＡＣとＤＮＡ配列５０２中のＤＮＡ４文字ＡＴＣＴを比較している。ＤＮＡ配列５０１に対応する中間ＤＮＡ配列中のコドンをＡＡＣ、ＤＮＡ配列５０２に対応する中間ＤＮＡ配列中のコドンをＴＣＴとしたときのアーク２０６のスコアは、ＤＮＡ配列５０１中のＤＮＡ３文字ＡＡＣとＤＮＡ配列５０１に対応する中間ＤＮＡ配列中のコドンＡＡＣとの間のスコアと、ＤＮＡ配列５０２中のＤＮＡ４文字ＡＴＣＴとＤＮＡ配列５０２に対応する中間ＤＮＡ配列中のコドンＴＣＴとの間のスコアと、ＤＮＡ配列５０１に対応する中間ＤＮＡ配列中のコドンＡＡＣを翻訳したアミノ酸ＡｓｎとＤＮＡ配列５０２に対応する中間ＤＮＡ配列中のコドンＴＣＴを翻訳したアミノ酸Ｓｅｒとの間のスコアの所定の関数の値となる。そしてアーク２０６のスコアは、ＤＮＡ配列５０１に対応する中間ＤＮＡ配列とＤＮＡ配列５０２に対応する中間ＤＮＡ配列に対して全ての塩基の組合せでできるコドンに対して求めた前記の所定の関数の最大値である。

【００３９】中間ＤＮＡ配列中のコドンは３文字だが、中間ＤＮＡ配列中のコドンと比較するＤＮＡ配列中のＤＮＡは３文字である必要はない。アーク２０６はその一例である。図４に示した翻訳したアミノ酸の比較に対応するアークは一例であり、ノード１０１の上に位置するノード１０２、１０３、１０４などからノード１０１へのアークすべてが翻訳したアミノ酸の挿入に対応するアークであり、ノード１０１の左に位置するノード１０５、１０６、１０７、１０８などからノード１０１へのアークすべてが翻訳したアミノ酸の欠失に対応するアークであり、ノード１０１の左上に位置するノード１０９、１１０、１１１などからノード１０１へのアークすべてが翻訳したアミノ酸どうしの比較に対応するアークとなる。これらすべてのアークに対して、アークの始まりのノードのスコアとアークのスコアの和を求め、最大値を求めることでノード１０１のスコアが求められる。

【００４０】図４に示した中間ＤＮＡ配列を用いてアミノ酸翻訳して比較するダイナミックプログラミング法による計算は、全ての中間ＤＮＡ配列を考慮することによってＤＮＡ配列測定における測定誤りを考慮しており、ＤＮＡ配列の途中にＤＮＡ挿入が入ってコドンの読み枠がずれた場合にも対応できる。

【００４１】ダイナミックプログラミング法では計算時間が長く問題となる場合には、生成する中間ＤＮＡ配列を制限することも可能である。翻訳方向と読み枠を固定した中間ＤＮＡ配列４０３として、ＤＮＡ配列４０１を順方向と逆方向にそれぞれ３種類の読み枠で固定したＤＮＡ配列を考える。中間ＤＮＡ配列４０４も同様に６種類考える。この場合中間ＤＮＡ配列４０３と中間ＤＮＡ配列４０４の組み合せは３６通りであり、全ての組み合せでスコア計算を行っても計算時間は長くない。

【００４２】図５は、本発明の一装置構成を示す図である。図５を用いて、計算処理手順を示す。スコア表記憶器１には上記方法で計算したデータベース内の各ＤＮＡ配列と一定長のあらゆるＤＮＡ配列とのアミノ酸配列に翻訳した時の類似性を示すスコアを、スコア表として作成しておく。検索の際は始めにデータベース各配列の近似スコアを記憶しておく近似スコア記憶器２と一時スコア記憶器８をすべて０にリセットし、検索配列を検索配

列記憶器３に記憶する。

【００４３】次に、カウンタ４を０から増分させていく。カウンタ４の出力は上位桁は部分配列取り出し器５に入力され部分配列切り出し部分を指定し、下位桁はスコア表記憶器１と近似スコア記憶器２と一時スコア記憶器８に入力されデータベース配列番号を指定する。指定された各配列の近似スコアと一時スコアはそれぞれスコア記憶器１４と１５に記憶される。検索配列記憶器３に接続された部分配列取り出し器５によって検索配列は一定長ｋ文字づつ切り出され、カウンタ４によって指定された部分配列が部分配列記憶器６に記憶される。部分配列記憶器６の出力はスコア表記憶器１に入力されカウンタ４で指定されたデータベース内の各配列とのダイナミックプログラミング法によるスコアがスコア表記憶器１から出力される。スコア表記憶器１の出力は加算器７を用いてスコア記憶器１４のスコアと加算される。加算器７の出力は比較器９で０と比較され大きい値が出力され、スコア記憶器１４に書き戻される。比較器９の出力は、スコア記憶器１５のスコアと比較器１０で比較され大きい値が再びスコア記憶器１５に書き戻される。その後スコア記憶器１４と１５の内容はそれぞれカウンタ４で指定される一時スコア記憶器８と近似スコア記憶器２の対応する部分に書き戻される。

【００４４】カウンタ４により検索配列から切り出されたすべての部分配列に対してすべてのデータベース配列が走査しつくされたら、近似スコア記憶器２の各スコアをソーター１１によって大きいスコアの順に並べ変え、配列名称記憶器１３によって対応する配列名称を求め表示器１２に表示する。次に表示されたＤＮＡ配列データベース中の配列の中から指定されたＤＮＡ配列と検索ＤＮＡ配列を、アミノ酸配列に翻訳して比較する。

【００４５】図３の翻訳比較では、ＤＮＡ配列と中間ＤＮＡ配列との並置、中間ＤＮＡ配列のアミノ酸配列への翻訳、翻訳されたアミノ酸配列どうしの並置を全て表示している。従来のアミノ酸配列どうしの並置と同様にアミノ酸配列の類似部分を表示することによって、共通する蛋白構造の位置、配列、類似度を同時に示している。これは蛋白への薬物結合サイトの発見から薬物設計、そして蛋白の共通機能の推定などに役立てることができる。同時に表示される中間ＤＮＡ配列のアミノ酸配列への翻訳からは用いられているコドンを調べることができる。生物種によって用いられるコドンにはかたよりがあるため、生物種がわかっていればここで表示されている翻訳が正しいかどうかを推定することができる。またＤＮＡ配列と中間ＤＮＡ配列との並置からはＤＮＡ配列測定誤差の傾向を知ることができる。これはＤＮＡ配列測定実験の精度を上げるために役立てられ、測定誤差の大きい部分のＤＮＡ配列を再測定するための指針となる。

【００４６】本実施例ではＤＮＡ配列と中間ＤＮＡ配列との並置、中間ＤＮＡ配列のアミノ酸配列への翻訳、翻訳されたアミノ酸配列どうしの並置を同時に表示することで、ＤＮＡ配列測定誤差、アミノ酸翻訳誤り、アミノ酸配列共通構造を一目で総合的に判断できる。本実施例と異なり中間ＤＮＡ配列、ＤＮＡ配列と中間ＤＮＡ配列との並置を省略することも可能である。省略した場合より多くの結果を表示することが可能となる。さらに全くＤＮＡ配列の表示がなく翻訳されたアミノ酸配列の並置だけを表示した場合にもＤＮＡ配列だけの情報から類似アミノ酸配列を知ることができるため有効である。また複数種類のフォントまたは色が表示可能な表示器の場合には、ＤＮＡ配列とアミノ酸配列のフォントまたは色に異なるものを用いることが可能である。これによりアミノ酸配列をＤＮＡ配列よりも強調し、ＤＮＡ配列の比較表示の中でもっとも重要であるアミノ酸配列の並置が一目でわかるようにすることができる。

【００４７】本実施例で用いている近似スコアは検索ＤＮＡ配列を分割した部分ＤＮＡ配列とＤＮＡ配列データベース中のＤＮＡ配列とのアミノ酸翻訳をして比較したスコアを加算した結果となっている。本実施例ではあらかじめデータベース内の各配列に対して一定長のあらゆる文字列とのアミノ酸翻訳をして比較したスコア表を作成してある。そのため近似スコアはスコア表を参照し和をとるだけで求めることが出来、計算量は小さい。

【００４８】本実施例では２つのＤＮＡ配列間でアミノ酸に翻訳して比較しているが、本発明は３つ以上のＤＮＡ配列間にも適用可能である。多数のＤＮＡ配列を本方法で比較することによって翻訳されたアミノ酸配列に共通する構造を見つけ出すことができる。

【００４９】

【発明の効果】本発明により、ＤＮＡ配列に対して翻訳されたアミノ酸配列構造がわからない場合にもアミノ酸配列の共通構造を発見できる。アミノ酸配列の共通構造の発見は、薬物の結合サイトの発見から薬物設計へ応用したり、類似蛋白機能の発見から蛋白設計などに役立つ。

【００５０】また、ＤＮＡ配列のアミノ酸配列への翻訳とアミノ酸配列どうしの並置表示を同時に行うことでＤＮＡ配列測定誤差、アミノ酸翻訳誤り、アミノ酸配列共通構造を一目で総合的に判断できる。

【００５１】また、ＤＮＡ配列から翻訳されるアミノ酸配列を比較することで、翻訳されるアミノ酸が互いに類似するＤＮＡ配列をもれなく検索することができる。

【図面の簡単な説明】

【図１】ＤＮＡ配列データベース内のＤＮＡ配列と固定長ＤＮＡ配列とをアミノ酸配列に翻訳してスコアを計算する方法を説明する図。

【図２】選択されたＤＮＡ配列と検索ＤＮＡ配列とをアミノ酸配列に翻訳してスコアを求める方法を説明する図。

【図３】実施例の表示画面を示す図。

15

【図４】中間ＤＮＡ配列をアミノ酸に翻訳して比較するときの、ダイナミックプログラミング法によるスコア計算方法を説明する図。

【図５】本発明の一装置構成を示す図。

【図６】アミノ酸翻訳コドン表を示す図。

【図７】ダイナミックプログラミング法を説明する図。

【図８】アミノ酸スコア表を示す図。

【符号の説明】

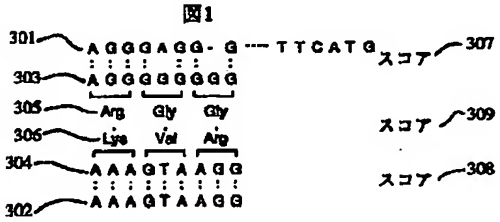
１…スコア表記憶器、２…近似スコア記憶器、３…検索

16

配列記憶器、４…カウンタ、５…部分配列取り出し器、６…部分配列記憶器、７…加算器、８…一時スコア記憶器、９，１０…比較器、１１…ソーター、１２…表示器、１３…配列名称記憶器、１４，１５…スコア記憶器、１０１～１１３…ノード、２０１～２０６…アーク、３０１，３０２，４０１，４０２，５０１，５０２…ＤＮＡ配列、３０３，３０４，４０３，４０４…中間ＤＮＡ配列、３０５，３０６，４０５，４０６…アミノ酸配列、３０７～３０９，４０７～４０９…スコア。

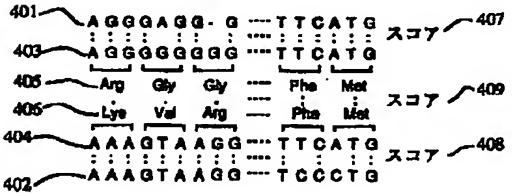
【図１】

図1



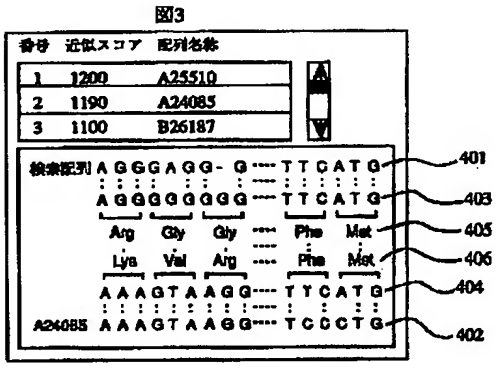
【図２】

図2



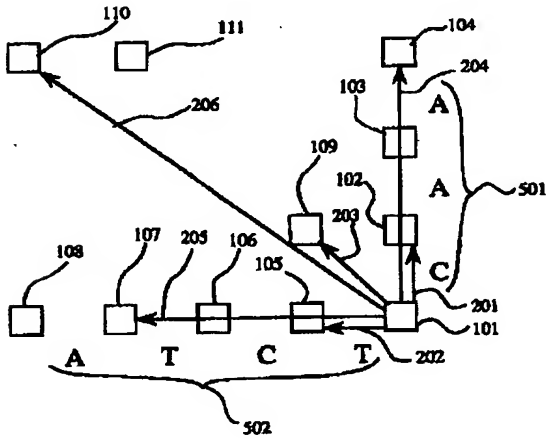
【図３】

図3



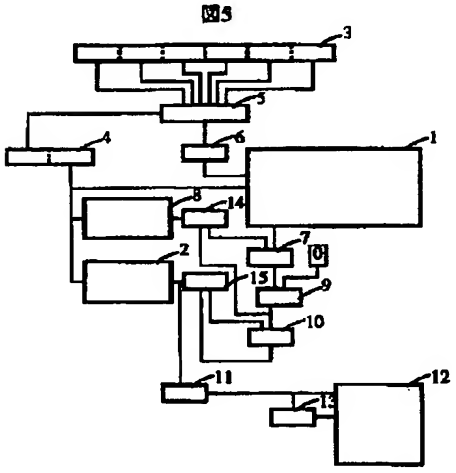
【図４】

図4



【図５】

図5



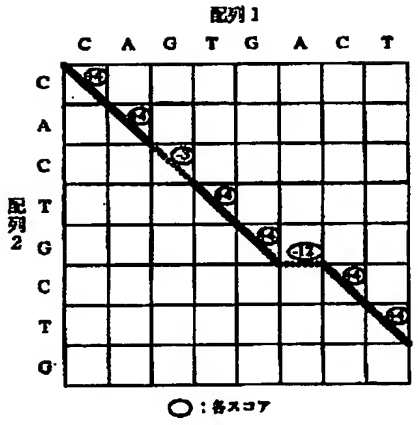
【図6】

図6

1番目 (5'末端)	2番目 U	2番目 C	2番目 A	2番目 G	3番目 (3'末端)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	終止	終止	A
	Leu	Ser	終止	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

【図7】

図7



【図8】

図8

システイン	C	12	イオウ結合性																						
セリン	S	0	2	小型																					
トレオニン	T	−2	1	3																					
プロリン	P	−3	1	0	6																				
アラニン	A	−2	1	1	1	2																			
グリシン	G	−3	1	0	−1	1	5																		
アスパラギン	N	−4	1	0	−1	0	0	2	酸性とそのアミド																
アスパラギン酸	D	−5	0	0	−1	0	1	2	4																
グルタミン酸	E	−5	0	0	−1	0	0	1	3	4															
グルタミン	Q	−5	−1	−1	0	0	−1	1	2	2	4														
ヒスチジン	H	−3	−1	−1	0	−1	−2	2	1	1	3	6	塩基性												
アルギニン	R	−4	0	−1	0	−2	−3	0	−1	−1	1	2	5												
リシン	K	−5	0	0	−1	−1	−2	1	0	0	1	0	3	5											
メチオニン	M	−5	−2	−1	−2	−1	−3	−2	−3	−2	−1	−2	0	0	6	疎水性									
イソロイシン	I	−2	−1	0	−2	−1	−3	−2	−2	−2	−2	−2	−2	−2	2	5									
ロイシン	L	−6	−3	−2	−3	−2	−4	−3	−4	−3	−2	−2	−3	−3	4	2	6								
バリン	V	−2	−1	0	−1	0	−1	−2	−2	−2	−2	−2	−2	−2	2	4	2	4							
フェニルアラニン	F	−4	−3	−3	−3	−4	−5	−4	−6	−5	−5	−2	−4	−3	0	1	2	−1	9	芳香族					
チロシン	Y	0	−3	−3	−3	−3	−5	−2	−4	−4	−4	0	−4	−1	−2	−1	−1	−2	7	10					
トリプトファン	W	−8	−3	−5	−5	−5	−7	−4	−7	−7	−5	−3	2	−1	−4	−5	−2	−6	0	0	17				
		C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W				

フロントページの続き

(72)発明者　笠原　直子
東京都国分寺市東恋ケ窪1丁目280番地
株式会社日立製作所中央研究所内